Defensive Publication: WhisperMic — Mic-side Pre-roll + Low-rate VAD Markers (+ Optional Subaudible Silence Flag)

Publication Date: August 13, 2025

Author: Liam Fahey (Public Domain / CC0 1.0)

Abstract

This publication describes a device-edge method for improving speech onset capture and reducing silence ingress into software and cloud AI pipelines. The method combines (1) a microphone/DSP pre-roll buffer (~100–500 ms) to prevent clipped first syllables upon unmute/PTT/VOX triggers; (2) a low-rate sideband of non-content markers (e.g., 20–50 Hz over HID/vendor-specific USB endpoint) indicating non-silence and optionally coarse event classes (impact, broadband noise); and (3) an optional subaudible control-tone fallback emitted only during confirmed silence for audio-only paths that cannot ingest metadata. The disclosure is intended to constitute prior art preventing exclusive patent claims over the general concept.

Enabling Disclosure

Pre-roll Buffer: The microphone’s MCU/DSP maintains a circular PCM buffer sized to 100–500 ms (user-selectable). On trigger (unmute/PTT/VOX), the device prepends the buffered audio to the outgoing stream, ensuring the first syllable is included. Cross-fade the splice with a 5–15 ms window to avoid clicks. Provide G HUB controls: enable toggle and duration slider.

Low-rate VAD Markers: A concurrent low-rate stream (e.g.,
20–50 Hz) is exposed via USB HID or a vendor-specific endpoint carrying booleans/labels: VAD (0=silence, 1=non-silence) and optional coarse events (impact/broadband). Markers are produced by an ultra-light DSP VAD using RMS with hysteresis and zero-crossing/spectral-flux gates. Apps/DAWs can ingest markers to suppress silent frames or auto-mark timelines; no audio content is included.

Subaudible Silence Flag (Fallback): For
audio-only chains lacking metadata, during confirmed silence the device may inject a very low-level control tone (default-off). Two variants: LF ~17–20 Hz (may be removed by high-pass filters) or HF ~19 kHz (possible audibility on some gear). Use amplitude limits around −50 dBFS and smooth attack/release. Auto-disable below 48 kHz sample rate or if safety constraints fail. Provide user controls and clear help text.
Compatibility/Privacy: All processing occurs on-device; markers convey non-content metadata. The method is model-agnostic and compatible with existing USB audio class streams.

Intended Effect: Reduce clipped-onset
occurrences; lower the proportion of silent frames transmitted or analysed by ASR/diarisation; decrease human edit time and storage/egress without modifying models.

Legal Notice

Public-Domain Dedication and Patent Non-Assertion: The author dedicates this disclosure to the public domain (CC0 1.0) and covenants not to assert any present or future patent claims the author owns or controls against implementations of the described method. This document is a non-confidential defensive publication intended to constitute prior art as of its publication date.